



Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury

Matthew M. Churpek, MD, MPH, PhD; Kyle A. Carey, MPH; Dana P. Edelson, MD, MS; Tripti Singh, MBBS; Brad C. Astor, PhD; Emily R. Gilbert, MD; Christopher Winslow, MD; Nirav Shah, MD, MPH; Majid Afshar, MD, MSCR; Jay L. Koyner, MD

Abstract

IMPORTANCE Acute kidney injury (AKI) is associated with increased morbidity and mortality in hospitalized patients. Current methods to identify patients at high risk of AKI are limited, and few prediction models have been externally validated.

OBJECTIVE To internally and externally validate a machine learning risk score to detect AKI in hospitalized patients.

DESIGN, SETTING, AND PARTICIPANTS This diagnostic study included 495 971 adult hospital admissions at the University of Chicago (UC) from 2008 to 2016 (n = 48 463), at Loyola University Medical Center (LUMC) from 2007 to 2017 (n = 200 613), and at NorthShore University Health System (NUS) from 2006 to 2016 (n = 246 895) with serum creatinine (SCr) measurements. Patients with an SCr concentration at admission greater than 3.0 mg/dL, with a prior diagnostic code for chronic kidney disease stage 4 or higher, or who received kidney replacement therapy within 48 hours of admission were excluded. A simplified version of a previously published gradient boosted machine AKI prediction algorithm was used; it was validated internally among patients at UC and externally among patients at NUS and LUMC.

MAIN OUTCOMES AND MEASURES Prediction of Kidney Disease Improving Global Outcomes SCr-defined stage 2 AKI within a 48-hour interval was the primary outcome. Discrimination was assessed by the area under the receiver operating characteristic curve (AUC).

RESULTS The study included 495 971 adult admissions (mean [SD] age, 63 [18] years; 87 689 [17.7%] African American; and 266 866 [53.8%] women) across 3 health systems. The development of stage 2 or higher AKI occurred in 15 664 of 48 463 patients (3.4%) in the UC cohort, 5711 of 200 613 (2.8%) in the LUMC cohort, and 3499 of 246 895 (1.4%) in the NUS cohort. In the UC cohort, 332 patients (0.7%) required kidney replacement therapy compared with 672 patients (0.3%) in the LUMC cohort and 440 patients (0.2%) in the NUS cohort. The AUCs for predicting at least stage 2 AKI in the next 48 hours were 0.86 (95% CI, 0.86-0.86) in the UC cohort, 0.85 (95% CI, 0.84-0.85) in the LUMC cohort, and 0.86 (95% CI, 0.86-0.86) in the NUS cohort. The AUCs for receipt of kidney replacement therapy within 48 hours were 0.96 (95% CI, 0.96-0.96) in the UC cohort, 0.95 (95% CI, 0.94-0.95) in the LUMC cohort, and 0.95 (95% CI, 0.94-0.95) in the NUS cohort. In time-to-event analysis, a probability cutoff of at least 0.057 predicted the onset of stage 2 AKI a median (IQR) of 27 (6.5-93) hours before the eventual doubling in SCr concentrations in the UC cohort, 34.5 (19-85) hours in the NUS cohort, and 39 (19-108) hours in the LUMC cohort.

(continued)

Key Points

Question What is the accuracy of a single-center machine learning algorithm for predicting acute kidney injury (AKI) when internally and externally tested?

Findings In this multicenter diagnostic study of approximately 500 000 admissions from 6 hospitals in 3 health systems, the machine learning algorithm had similarly high discrimination in both internal and external validation cohorts. Alert thresholds fired nearly a day and a half before the event.

Meaning These findings demonstrate that the AKI algorithm is generalizable to patients in the center in which it was derived and to patients from other hospitals, suggesting that implementation could prompt early identification and therapy aimed at decreasing preventable AKI.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

CONCLUSIONS AND RELEVANCE In this study, the machine learning algorithm demonstrated excellent discrimination in both internal and external validation, supporting its generalizability and potential as a clinical decision support tool to improve AKI detection and outcomes.

JAMA Network Open. 2020;3(8):e2012892. doi:10.1001/jamanetworkopen.2020.12892

Introduction

Acute kidney injury (AKI) is a common clinical syndrome in hospitalized patients and is associated with increased morbidity, mortality, and cost of care.¹⁻³ Consensus criteria define AKI by either an increase in serum creatinine (SCr) concentration or a decrease in urine output.⁴ Biomarkers that detect AKI prior to these changes have been investigated for several years. However, to date, there has been limited large-scale validation and implementation of these tools. Detection of AKI prior to the changes in SCr concentration may provide a crucial window of opportunity to prevent further injury and allow clinicians to intervene in the hopes of improving patient outcomes.

While work on urinary and serum biomarkers of early AKI continues,⁵⁻⁷ several groups have reported on the accuracy of electronic health record–based risk scores that can identify AKI before changes in SCr concentration.⁸⁻¹⁴ The scope of these published algorithms has varied, with some focusing on only ward or intensive care unit (ICU) patients and others on postoperative AKI.⁸⁻¹⁴ Additionally, these algorithms range from rule-based, more parsimonious scores to complex, machine learning–based scores.^{8,10,13} However, regardless of the individual score, there has been limited external validation of these risk assessment tools. Our group has previously published a gradient boosted machine learning AKI prediction model for all hospitalized patients (ie, patients in the emergency department, ward, and ICU) using single-center data at the University of Chicago (UC).¹⁰ We subsequently simplified this risk score and clinically implemented the streamlined version to prompt early nephrology consultation as part of a single-center randomized controlled trial.¹⁵ In this study, we aim to both internally (at UC) and externally validate the simplified version of our AKI score using retrospective cohorts from independent health systems (Loyola University Medical Center [LUMC], Maywood, Illinois, and Northshore University Health System [NUS], Evanston, Illinois).

Methods

Study Population

We included 3 distinct adult (≥ 18 years) patient cohorts in this retrospective cohort study of prospectively collected data. All admitted adult patients at UC (an urban tertiary referral hospital) who were part of the validation cohort (2008 to 2016) in our previously published AKI algorithm development study¹⁰ were used for internal validation of the model. External validation was performed in adult patients admitted to LUMC (a suburban tertiary referral hospital) from 2007 to 2017 and NUS (a suburban 4-hospital health care network) from 2006 to 2016. Patients were excluded if they had no documented SCr concentration during their admission; had an initial admitting SCr concentration of at least 3.0 mg/dL (to convert to micromoles per liter, multiply by 88.4); had diagnosis codes for stage 4 or higher chronic kidney disease from any prior inpatient or outpatient encounter; developed Kidney Disease Improving Global Outcomes (KDIGO) stage 2 AKI (ie, SCr concentrations doubled) in a location other than the ward, emergency department, or ICU; or required kidney replacement therapy (KRT) within 48 hours of their first documented SCr measurement.¹⁰ The study protocol was approved by the UC, LUMC, and NUS institutional review boards with a waiver of informed consent based on minimal harm and impracticability. We followed

the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.¹⁶

Data Collection

Demographic characteristics, patient location data (eg, ward, emergency department, ICU, operating room), vital signs, laboratory values, and nurse documentation were accessed through the Clinical Research Data Warehouse at UC. A similar process was completed at LUMC and NUS through their respective data warehouses, and data from the external sites were transferred to UC for analysis under a data use agreement.

AKI Definitions

We defined AKI by the SCr-based criteria from the KDIGO consensus definition.⁴ Baseline SCr concentration was defined as the admission SCr value and was updated on a rolling basis for 48-hour and 7-day criteria, as per the KDIGO guidelines.^{4,9,10}

Statistical Analysis

Patient characteristics, laboratory values, and outcomes were compared among the 3 cohorts (NUS, LUMC, and UC). These same factors were compared within the individual cohorts between patients who developed AKI and those who did not. We used *t* tests, Wilcoxon rank sum tests, analyses of variance, Kruskal-Wallis tests, and χ^2 tests for these comparisons, as appropriate, based on the distributions of the variables.

Next, the simplified version of our previously developed gradient boosted machine model, which was derived only using UC data, was applied to the UC internal validation cohort and the LUMC and NUS external validation cohorts. As previously described, the originally published gradient boosted machine model was developed using discrete time survival analysis, included 97 variables, and was developed and validated solely using UC data.¹⁰ This model was simplified to 59 variables, with model development performed as described in the prior publication¹⁰ using the same derivation cohort, with 10-fold cross-validation in the derivation data used to tune the model hyperparameters. Predictors in the simplified model include demographic characteristics, vital signs, routine chemistry and hematology laboratory values, trends of vital sign and laboratory values (eg, highest heart rate in previous 24 hours), and nursing documentation (eg, Braden score) (eTable 1 in the [Supplement](#)). Missing data were handled as previously described, with the median (for continuous data) or mode (for categorical data) by location being imputed for missing predictor values that remained after carry-forward imputation.¹⁰ eFigure 1 in the [Supplement](#) illustrates the variable importance plot for the top 15 variables in the simplified model developed from UC data. Of note, this is the same simplified model that is now running prospectively at UC as part of an National Institutes of Health-funded clinical trial.¹⁵ The simplified model, which was developed only in the previously described UC derivation cohort, was used to produce predicted probabilities for every new observation (eg, new vital sign or laboratory value) in the 3 validation cohorts. These probabilities were used to calculate the area under the receiver operating characteristic curve (AUC) using the trapezoidal method, with the Delong method for confidence intervals.¹⁷ For accuracy calculations, probabilities were calculated for every observation until the event of interest occurred or the patient was discharged.

The primary outcome of this study was the development of SCr-based stage 2 AKI within 48 hours of each observation. Accuracy metrics at individual probability thresholds were also calculated using the maximum score during the admission prior to the outcome of interest or discharge. Secondary outcomes included the development of stage 1 AKI, stage 3 AKI, receipt of KRT, and inpatient mortality. Subgroup analyses were performed by hospital location (ICU vs ward), admission SCr concentration strata, and time in an operating room. All analyses were performed using Stata version 15.1 (StataCorp) and R version 3.6.1 (The R Project for Statistical Computing). Statistical significance was set at $P < .05$, and all tests were 2-tailed.

Results

The final cohort included 495 971 adult patient admissions (mean [SD] age, 63 [18] years; 87 689 [17.7%] African American; and 266 866 [53.8%] women) across 6 hospitals at 3 health systems. Exclusions leading to this cohort were previously published for UC¹⁰ and are found in eFigure 2 in the Supplement for LUMC and NUS. Compared with the other 2 cohorts, patient admissions from UC were more likely to be younger (mean [SD] age: LUMC, 58.6 [17.2] years; NUS, 67.4 [17.7] years; UC, 56.6 [17.8] years; $P < .001$) and African American patients (LUMC, 45 512 [22.7%]; 17 940 [7.3%]; UC, 24 237 [50.0%]; $P < .001$) (Table 1). While statistically significant differences in the admission SCr and blood urea nitrogen (BUN) concentrations were seen across cohorts, the numerical differences were small. The UC internal validation cohort included 48 463 patient admissions, 6935 (14.3%) of whom developed at least stage 1 AKI, 1664 (3.4%) who developed stage 2 or 3 AKI, and 332 (0.7%) requiring KRT. Among the 200 613 patients included in the LUMC cohort, 27 352 (13.6%) developed at least stage 1 AKI, 5722 (2.8%) developed stage 2 or 3 AKI, and 672 (0.3%) required KRT. Among the 246 895 patients in the NUS cohort, 20 473 (8.3%) developed any AKI, 3499 (1.4%) developed stage 2 or 3 AKI, and 440 (0.2%) received KRT. eFigure 3 in the Supplement provides the cumulative incidence plots for all individual stages of AKI and receipt of KRT over time across all 3 cohorts.

eTable 2 in the Supplement provides the demographic characteristics and outcome data for all 3 cohorts, stratified by those with and without AKI. Compared with patients who did not develop AKI, those who did were more likely to be older (eg, mean [SD] age in NUS cohort, 66.8 [17.9] years vs 73.2 [14.7] years; $P < .001$); male (96 711 of 226 442 [42.7%] vs 10 226 of 20 473 [49.9%]; $P < .001$), and have higher mean (SD) admission SCr (1 [0.4] mg/dL vs 1.2 [0.6] mg/dL; $P < .001$) and BUN values (19 [12.5] mg/dL vs 26.1 [16.8] mg/dL [to convert to millimoles per liter, multiply by 0.357]; $P < .001$). Compared with patients who did not develop AKI, those who did were more likely to have been in an operating room (eg, LUMC cohort: 49 875 of 173 261 [28.8%] vs 11 938 of 27 352 [43.6%]; $P < .001$) or ICU (34 009 [19.6%] vs 15 794 [57.7%]; $P < .001$), had significantly longer median (interquartile range) hospital lengths of stay (2.4 [1.2-4.7] vs 7.9 [4.4-14.9]; $P < .001$), and had higher inpatient mortality (1403 [0.8%] vs 2883 [10.5%]; $P < .001$).

Table 1. Clinical Demographics and Outcomes of the 3 Patient Cohorts With 495 971 Total Patients

Characteristic	Cohort, No. (%)			P value
	LUMC (n = 200 613)	NUS (n = 246 895)	UC (n = 48 463)	
Age, mean (SD), y	58.6 (17.2)	67.4 (17.7)	56.6 (17.8)	<.001
African American	45 512 (22.7)	17 940 (7.3)	24 237 (50)	<.001
Female	100 987 (50.3)	139 947 (56.7)	25 932 (53.5)	<.001
Admission serum creatinine, mean (SD), mg/dL	1.1 (0.4)	1.0 (0.4)	1.0 (0.5)	<.001
Admission blood urea nitrogen, mean (SD), mg/dL	16.9 (11.8)	19.6 (13)	17.9 (12.1)	<.001
Patients developing any AKI				
Stage 1	27 352 (13.6)	20 473 (8.3)	6935 (14.3)	
Stage 2	5711 (2.8)	3499 (1.4)	1664 (3.4)	<.001
Stage 3	2575 (1.3)	1406 (0.6)	775 (1.6)	
Receipt of dialysis more than 48 h after initial serum creatinine measurement	672 (0.3)	440 (0.2)	332 (0.7)	<.001
Length of hospital stay, median (IQR), d	2.8 (1.3-5.8)	3 (1.7-4.8)	3.9 (2.1-6.7)	<.001
Location of AKI				
Ward	16 007 (8.0)	14 540 (5.9)	4818 (9.9)	
ICU	10 434 (5.2)	4735 (1.9)	1949 (4.0)	<.001
Emergency department or other	911 (0.5)	1198 (0.5)	168 (0.3)	
ICU admission during stay	49 803 (24.8)	33 869 (13.7)	10 059 (20.8)	<.001
Operating room during stay	61 813 (30.8)	61 940 (25.1)	13 442 (27.7)	<.001
Inpatient mortality	4286 (2.1)	3215 (1.3)	1050 (2.2)	<.001

Abbreviations: AKI, acute kidney injury; ICU, intensive care unit; IQR, interquartile range; LUMC, Loyola University Medical Center; NUS, Northshore University Health System; UC, The University of Chicago.

SI conversion factors: To convert blood urea nitrogen to millimoles per liter, multiply by 0.357; serum creatinine to micromoles per liter, multiply 88.4.

Model discrimination results for the prediction of all stages of AKI and the need for KRT in the next 48 hours across all 3 cohorts are shown in **Table 2**. The AUCs were the same or slightly higher in the UC cohort for all outcomes. The model predicted the development of stage 2 AKI within 48 hours with an AUC of 0.86 (95% CI, 0.86-0.86) in the UC cohort, 0.86 (95% CI, 0.86-0.86) in the NUS cohort, and 0.85 (95% CI, 0.84-0.85) in the LUMC cohort. The model provided excellent discrimination of those needing KRT within 48 hours, with AUCs of 0.95 or higher in all 3 cohorts. **Table 3** provides the AUCs for the prediction of stage 2 AKI in the next 48 hours across all 3 cohorts stratified by patient location, admission SCr concentration, and prior operating room status. In the UC and LUMC cohorts, the model had slightly higher discrimination for the prediction of stage 2 AKI for patients in the ICU compared with ward patients, although these differences were small (difference, 0.01-0.02). The model had very similar discrimination for the prediction of stage 2 AKI in the next 48 hours on the wards in all 3 cohorts. In all 3 cohorts, the model performed better among patients with higher admission SCr concentrations, performing best among those with an admission SCr concentration between 2.0 and 2.9 mg/dL. In all subgroups across all sites, the AUC for the development of stage 2 AKI in the next 48 hours was greater than 0.80. eTable 3 in the [Supplement](#) provides the AUCs for the prediction of stage 2 AKI in the next 24 hours across all 3 cohorts in the same subgroups, which were numerically higher than the results for predicting events within 48 hours. For example, for patients in the NUS with admission SCr concentrations of 2.0 to 2.9, the model had an AUC of 0.93 (95% CI, 0.92-0.93). Subgroup analyses looking at postoperative patients demonstrated that the algorithm performed nearly identically among those who did and did not previously go to an operating room (Table 3; eTable 3 in the [Supplement](#)). Table 3 also shows that, in all 3 cohorts, the model provided an AUC of greater than 0.84 for the prediction of postoperative stage 2 AKI in the next 48 hours. Calibration plots for all 3 sites are shown in eFigure 4 in the [Supplement](#), which mimics how the model was derived using 12-hour blocked data. The model was

Table 2. AUC for the Development of AKI and the Receipt of KRT Within the Next 48 Hours^a

Outcome	AUC (95% CI), by cohort		
	LUMC (n = 200 613)	NUS (n = 246 895)	UC (n = 48 463)
AKI in 48 h			
≥Stage 1	0.67 (0.67-0.67)	0.69 (0.69-0.69)	0.72 (0.71-0.72)
≥Stage 2	0.85 (0.84-0.85)	0.86 (0.86-0.86)	0.86 (0.86-0.86)
≥Stage 3	0.91 (0.91-0.91)	0.92 (0.92-0.92)	0.92 (0.92-0.92)
Receipt of KRT			
In 24 h	0.96 (0.96-0.96)	0.96 (0.95-0.96)	0.97 (0.97-0.97)
In 48 h	0.95 (0.95-0.95)	0.95 (0.94-0.95)	0.96 (0.96-0.96)
In 72 h	0.94 (0.94-0.94)	0.93 (0.93-0.93)	0.94 (0.94-0.94)

Abbreviations: AKI, acute kidney injury; AUC, area under the receiver operating characteristic curve; LUMC, Loyola University Medical Center; NUS, Northshore University Health System; KRT, kidney replacement therapy; UC, The University of Chicago.

^a A total of 27 532, 20 473, and 6935 admissions developed stage 1 AKI, 5711, 3499, and 1664 developed stage 2 AKI, 2575, 1406, and 775 developed stage 3 AKI, and 672, 440, and 332 received KRT at LUMC, NUS, and UC, respectively.

Table 3. AUCs for the Model to Predict Stage 2 AKI in the Next 48 Hours in All Cohorts Stratified by Patient Location, Admission Serum Creatinine Level, and Time in Operating Room

Variable	AUC for predicting stage 2 AKI within 48 h (95%CI), by cohort		
	LUMC (n = 200 613)	NUS (n = 246 895)	UC (n = 48 463)
Patient location			
Ward	0.80 (0.80-0.80)	0.84 (0.84-0.84)	0.83 (0.83-0.83)
Intensive care unit	0.82 (0.82-0.82)	0.83 (0.83-0.83)	0.84 (0.84-0.85)
Admission serum creatinine, mg/dL			
<1.0	0.81 (0.81-0.82)	0.83 (0.83-0.83)	0.84 (0.84-0.84)
1.0 to <2.0	0.86 (0.86-0.86)	0.88 (0.88-0.88)	0.88 (0.88-0.88)
2.0-2.9	0.89 (0.88-0.89)	0.89 (0.88-0.89)	0.89 (0.89-0.90)
Time spent in an operating room			
Prior operating room	0.84 (0.84-0.85)	0.86 (0.86-0.86)	0.86 (0.86-0.86)
No prior operating room	0.85 (0.85-0.85)	0.86 (0.86-0.86)	0.86 (0.86-0.86)

Abbreviations: AKI, acute kidney injury; AUC, area under the receiver operating characteristic curve; LUMC, Loyola University Medical Center; NUS, Northshore University Health System; UC, The University of Chicago.

SI conversion factor: To convert serum creatinine to micromoles per liter, multiply by 88.4.

well calibrated at all sites, except for the highest risk decile at UC and the top 2 risk deciles at LUMC and NUS.

Table 4 demonstrates the sensitivity, specificity, and positive and negative predictive values (PPV and NPV) for each probability cutoff using the maximum score for each admission to predict stage 2 AKI during the admission. Several probability cutoff values provided high sensitivity and specificity, with a cutoff of at least 0.057 providing a sensitivity of 87.1%, an NPV of 99.5%, and a PPV of 27.0% in the UC cohort. Similar or slightly lower accuracy results were seen in the LUMC and NUS cohorts across different thresholds (Table 4). eTable 4 in the [Supplement](#) demonstrates the same

Table 4. Accuracy and Timing of Detection of Different Probability Cutoffs for Detecting Stage 2 AKI Using the Maximum Score During the Admission Prior to the Event or Discharge

Probability cutoff	Patients, No. ^a	Time to stage 2, median (IQR), h	Sensitivity, %	Specificity, %	PPV	NPV
LUMC cohort						
≥0.010	50 860	64 (25-175)	90.1	76.5	10.1	99.6
≥0.030	30 580	49 (23-138)	81.7	86.7	15.2	99.4
≥0.045	24 083	44 (21-120)	76.4	89.9	18.1	99.2
≥0.057	20 520	39 (19-108)	73.3	91.6	20.4	99.2
≥0.075	16 508	35 (17-97)	68.0	93.5	23.5	99.0
≥0.100	12 539	28 (15-83)	60.8	95.3	27.6	98.8
≥0.125	9533	25 (13-71)	53.1	96.7	31.7	98.6
≥0.150	7288	24 (11-64)	45.5	97.6	35.6	98.4
≥0.175	5406	23 (11-55)	37.3	98.3	39.3	98.2
≥0.200	3945	22 (9-48)	29.3	98.8	42.4	98.0
≥0.250	2001	18 (8-35)	17.1	99.5	48.6	97.6
≥0.400	52	11 (4-27)	0.6	100.0	63.5	97.2
NUS cohort						
≥0.010	44 439	53 (24-133)	86.1	83.0	6.7	99.8
≥0.030	24 048	44 (22-106)	75.2	91.2	10.8	99.6
≥0.045	18 165	39 (21-95)	69.3	93.5	13.2	99.5
≥0.057	15 137	34.5 (19-85)	64.4	94.7	14.7	99.5
≥0.075	11 748	30 (18-73)	58.3	96.0	17.1	99.4
≥0.100	8632	26 (16-65)	50.9	97.2	20.4	99.3
≥0.125	6339	25 (14-54)	42.9	98.0	23.4	99.2
≥0.150	4675	24 (13-48)	36.5	98.6	26.9	99.1
≥0.175	3375	23 (11-46)	29.1	99.0	29.8	99.0
≥0.200	2374	21 (10-41)	22.8	99.3	33.2	98.9
≥0.250	1066	18 (8-32)	11.8	99.7	38.4	98.8
≥0.400	11	21 (8-34)	0.1	100.0	18.2	98.6
UC cohort						
≥0.010	13 756	57.5 (22-183.5)	99.1	74.1	12.0	100.0
≥0.030	7971	38 (11-130)	94.3	86.3	19.6	99.8
≥0.045	6249	31 (9-107)	90.6	89.9	24.1	99.6
≥0.057	5360	27 (6.5-93)	87.1	91.6	27.0	99.5
≥0.075	4379	24 (4-74)	81.4	93.5	30.9	99.3
≥0.100	3367	22 (2-56)	73.0	95.4	36.0	99.0
≥0.125	2627	18 (0-47)	64.8	96.7	41.0	98.7
≥0.150	2054	16 (0-37)	56.4	97.6	45.6	98.4
≥0.175	1561	14 (0-30)	45.8	98.3	48.8	98.1
≥0.200	1177	12 (0-27)	35.6	98.8	50.3	97.7
≥0.250	605	10 (0-25)	20.9	99.4	57.4	97.3
≥0.400	7	18 (2-23)	0.3	100.0	71.4	96.6

Abbreviations: AKI, acute kidney injury; IQR, interquartile range; LUMC, Loyola University Medical Center; NPV, negative predictive value; NUS, Northshore University Health System; PPV, positive predictive value; UC, The University of Chicago.

^a Number of patients denotes the number of encounters to reach the threshold before first meeting stage 2 AKI.

performance metrics using every observation in the test data sets for whether the outcome occurred within 48 hours at each individual probability cutoff across all 3 cohorts.

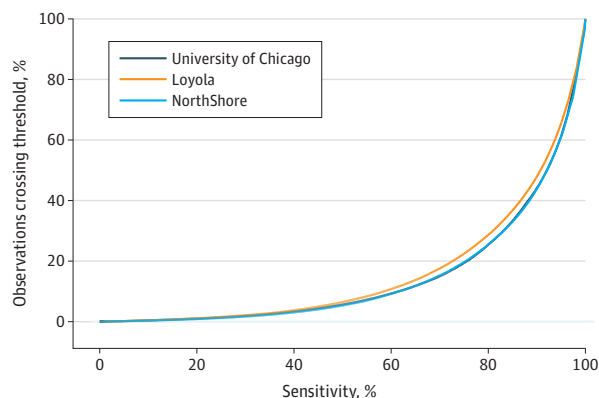
The utility of the model as a decision support tool, with an illustration of the percentage of observations that crossed each alert threshold by the sensitivity of that threshold for predicting the development of stage 2 AKI within 48 hours, is shown in the **Figure**. As shown, relatively fewer alerts would fire at UC and NUS compared with LUMC if a high ($\geq 60\%$) sensitivity was desired. In a time-to-event analysis, a cutoff of at least 0.057 predicted the later onset of stage 2 AKI a median (IQR) of 27 (6.5-93) hours before the eventual doubling in SCr concentration in the UC cohort, 34.5 (19-85) hours in the NUS cohort, and 39 (19-108) hours in the LUMC cohort. Table 4 provides time-to-event analysis for all cutoffs across all cohorts.

Discussion

In this large, multicenter study across 6 hospitals, 3 health systems, and nearly 500 000 patient admissions, we performed an internal and external validation of a machine learning risk algorithm that predicts the development of AKI across all hospitalized patients. Our findings demonstrate consistent, high discrimination across all sites, hospital locations, and baseline SCr values as well as higher discrimination for the more severe forms of AKI (ie, stage 3 AKI and the need for KRT). Importantly, the model identified patients at risk of AKI nearly a day and a half earlier than the current criterion standard, ie, SCr concentration. This advanced notice could potentially allow for preemptive interventions for patients at high risk of AKI, which could improve outcomes. Our model, which has now been validated in 2 unique, external health systems, uses clinical data that is readily available in the electronic health record and can be implemented for real-time use.¹⁵

Although model accuracy often decreases during external validation, we found similar results for predicting severe AKI in the internal and external validation cohorts. This may be because AKI is defined using SCr concentrations, and our model included mostly generalizable physiologic variables. As expected, discrimination was slightly higher in the UC internal validation cohort in some subgroups, with the largest difference in performance seen in those with stage 1 AKI. Because this stage of AKI can be affected by benign fluid shifts and fluid administration practices and may not represent true kidney tubular injury, this is likely a less important outcome to predict than more severe stages of AKI.¹⁸⁻²⁰ Our model's strengths include its ability to detect AKI in those with and without an elevated SCr concentration at admission. The model demonstrated AUCs greater than 0.90 in the validation cohorts for detecting stage 2 AKI in the next 24 hours in those with an admission SCr concentration greater than 1.0 mg/dL. This discrimination was slightly lower (≥ 0.85 AUC) for predicting stage 2 AKI in the next 48 hours, regardless of admission SCr concentration.

Figure. Efficiency Curves for Predicting Stage 2 AKI for All Included Cohorts



Other groups have developed electronic risk scores for the detection of AKI. For example, Wilson and colleagues⁸ developed a parsimonious model using retrospective data from 169 859 hospitalized adults across 3 hospitals in the same health care system. They demonstrated that a simple model based on available laboratory data can accurately predict stage 1 AKI with an AUC of 0.74. Of note, this model was built and trained to predict the more common stage 1 AKI rather than the more severe stage 2, which we used as the primary outcome of our model. In addition, Lei and colleagues¹⁴ used a gradient boosted machine learning model to identify patients at risk of the development of postoperative AKI in a multicenter study. They demonstrated that through the addition of prehospitalization, preoperative, and intra-operative variables, they improved their ability to detect postoperative KDIGO SCr-based stage 1 AKI, with an AUC of 0.82 in their final model.¹⁴

Recently, Tomasev and colleagues¹³ published a risk score using data from more than 703 000 adult patients in the United States Department of Veterans Affairs Healthcare System. Using a recurrent neural network, they developed an accurate model that could detect KDIGO-defined AKI. Their final model provided a sensitivity of 55.8% and specificity of 82.7%, based on a 2:1 false-to-true alert ratio. However, they used a randomly selected group of patients to serve as their test set, which has been shown to provide optimistic estimates of accuracy compared with external validation.²¹ Additionally, they used deep learning and included 620 000 features, which would be much less interpretable than our model, which had 59 features. Furthermore, the Veteran Affairs data set remains limited because it included only 6.4% female patients and has unknown validity in more diverse settings. In contrast, our cohort was nearly 50% women and included 17.7% African American patients. Furthermore, we included urban and suburban academic centers as well as community nonteaching hospitals, which increases the generalizability of our findings.

Once electronic scores like ours have been developed, their clinical utility needs to be investigated. Several recent studies have demonstrated that early nephrology care in the setting of increased AKI risk or early AKI is associated with improved patient outcomes. Selby and colleagues²² performed a multicenter, pragmatic, step-wedge cluster randomized trial across 5 hospitals in the United Kingdom involving the implementation of early AKI e-alerts as well as AKI care bundles and a kidney care-focused educational program for all hospitalized patients. They demonstrated improved quality of kidney care, with shorter length of stay, improved AKI recognition, medication optimization, and fluid assessment.²² Similarly, in 2 separate postoperative cohorts, the use of a KDIGO care bundle in patients identified as high risk of severe AKI (via elevation of their urinary tissue inhibitor metalloprotease-2 and insulin-like growth factor binding protein 2 concentration) led to improved patient outcomes.^{23,24} Meersch and colleagues²³ demonstrated a decrease in the incidence and severity of AKI with the use of a care bundle following cardiac surgery, while Gőcze and colleagues²⁴ demonstrated decreased AKI severity and reduction in ICU length of stay after major abdominal surgery.^{23,24} As such, it is reasonable to expect that the implementation of an early electronic AKI risk score may similarly improve AKI outcomes. However, these novel tools should be implemented and then thoroughly investigated to determine their utility.

Using a tool like ours would involve implementing an intervention the first time a patient reaches a unique risk threshold to augment our ability to prevent AKI. Table 4 demonstrates the test performance for the first time patients meet unique probability cutoffs. As shown with these results, there are several thresholds with adequate PPV and sensitivity values that could be used in clinical practice. Future work to determine the optimal threshold for clinical action that balances detection rates and false alarms, which will require interventional trials, is needed.

Limitations

Our study has limitations. We only defined AKI through changes in SCr concentration because of the inability to obtain accurate hourly urine output measurements in all hospitalized patients to comply with KDIGO definitions.⁴ However, this is in line with several other previously published AKI risk scores.^{8,13} Additionally, given the limitations of all 3 data sets (eg, only having access to inpatient data), we defined baseline SCr concentration using the admission values as opposed to outpatient

values. This is similar to how we developed our original models and is in line with current standards in AKI research.^{9,10,25} Future study is needed to determine how this assumption affects model accuracy. We excluded patients without any SCr measured during their admission, because whether they developed AKI was unknowable, so our model does not apply to these patients. Another limitation is that our model overpredicted risk for the highest decile of patients, as shown in the calibration plot (eFigure 4 in the Supplement). However, in clinical practice, our focus is in identifying the highest risk patients, so the ordering of patients (as it relates to discrimination) is more important to our clinical workflow than calibration. Finally, the external validation cohorts were mostly teaching hospitals, and all were in Illinois. However, the comparable results, diverse patient populations, and reliance on mostly physiological variables (as opposed to variables such as billing codes, which can vary across hospitals and over time), suggest that our model is likely to be generalizable to a number of other settings.

Conclusions

In this study, we internally and externally validated a novel machine learning risk score for the prediction of AKI across all hospital settings. This tool, which includes patient demographic characteristics, vital signs, laboratory values, and nursing assessments, can be used to identify patients at increased risk of the development of severe AKI and the need for KRT. Pairing this risk score with early, kidney-focused care may improve outcomes in the patients at the highest risk of the development of AKI.

ARTICLE INFORMATION

Accepted for Publication: May 28, 2020.

Published: August 11, 2020. doi:10.1001/jamanetworkopen.2020.12892

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2020 Churpek MM et al. *JAMA Network Open*.

Corresponding Author: Jay L. Koyner, MD, Department of Medicine, The University of Chicago, 5481 S Maryland Ave, Ste S-506, MC5100, Chicago, IL 60637 (jkoyner@uchicago.edu).

Author Affiliations: Department of Medicine, University of Wisconsin, Madison (Churpek, Singh, Astor); Department of Medicine, The University of Chicago, Illinois (Carey, Edelson, Shah, Koyner); Department of Population Health Sciences, University of Wisconsin, Madison (Astor); Department of Medicine, Loyola University Medical Center, Maywood, Illinois (Gilbert, Afshar); Department of Medicine, NorthShore University Healthcare, Evanston, Illinois (Winslow, Shah).

Author Contributions: Dr Churpek and Mr Carey had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Churpek, Edelson, Singh, Koyner.

Acquisition, analysis, or interpretation of data: Churpek, Carey, Astor, Gilbert, Winslow, Shah, Afshar, Koyner.

Drafting of the manuscript: Churpek, Carey, Edelson, Singh, Koyner.

Critical revision of the manuscript for important intellectual content: Churpek, Carey, Edelson, Astor, Gilbert, Winslow, Shah, Afshar.

Statistical analysis: Churpek, Carey.

Obtained funding: Koyner.

Administrative, technical, or material support: Churpek, Edelson, Winslow, Shah, Koyner.

Supervision: Edelson, Winslow, Koyner.

Conflict of Interest Disclosures: Dr Churpek reported receiving grants from EarlySense Research and the National Institute of General Medicine Sciences outside the submitted work and having a patent for risk stratification algorithms for hospitalized patients pending. Dr Edelson reported receiving grants from EarlySense Research, Philips Healthcare, the American Heart Association, and Laerdal Medical outside the submitted work; having a patent for risk stratification algorithms for hospitalized patients pending; being the president/cofounder and a

minority shareholder of AgileMD, which develops clinical decision support tools for hospitals; being the chair for the American Heart Association Get With the Guidelines adult research task force; and having an ownership interest in Quant HC, which is developing products for risk stratification of hospitalized patients. No other disclosures were reported.

Funding/Support: Drs Churpek, Edelson, and Koyner were supported by grant R21DK113420 from the National Institute of Diabetes and Digestive and Kidney Diseases. Drs Churpek, Edelson, Winslow, Shah, and Afshar and Mr Carey were supported by grant R01 GM123193 from the National Institute of General Medicine Sciences.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol*. 2005;16(11):3365-3370. doi:10.1681/ASN.2004090740
2. Hobson C, Ozrazgat-Baslanti T, Kuxhausen A, et al. Cost and mortality associated with postoperative acute kidney injury. *Ann Surg*. 2015;261(6):1207-1214. doi:10.1097/SLA.0000000000000732
3. Chertow GM, Levy EM, Hammermeister KE, Grover F, Daley J. Independent association between acute renal failure and mortality following cardiac surgery. *Am J Med*. 1998;104(4):343-348. doi:10.1016/S0002-9343(98)00058-8
4. KDIGO. KDIGO clinical practice guideline for acute kidney injury. Published March 2012. Accessed July 10, 2020. <https://kdigo.org/wp-content/uploads/2016/10/KDIGO-2012-AKI-Guideline-English.pdf>
5. Xie Y, Ankawi G, Yang B, et al. Tissue inhibitor metalloproteinase-2 (TIMP-2) • IGF-binding protein-7 (IGFBP7) levels are associated with adverse outcomes in patients in the intensive care unit with acute kidney injury. *Kidney Int*. 2019;95(6):1486-1493. doi:10.1016/j.kint.2019.01.020
6. Mansour SG, Zhang WR, Moledina DG, et al; TRIBE-AKI Consortium. The association of angiogenesis markers with acute kidney injury and mortality after cardiac surgery. *Am J Kidney Dis*. 2019;74(1):36-46. doi:10.1053/j.ajkd.2019.01.028
7. Koyner JL, Zarbock A, Basu RK, Ronco C. The impact of biomarkers of acute kidney injury on individual patient care. *Nephrol Dial Transplant*. 2019;gfgz188. doi:10.1093/ndt/gfz188
8. Simonov M, Ugwuowo U, Moreira E, et al. A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: a descriptive modeling study. *PLoS Med*. 2019;16(7):e1002861. doi:10.1371/journal.pmed.1002861
9. Koyner JL, Adhikari R, Edelson DP, Churpek MM. Development of a multicenter ward-based AKI prediction model. *Clin J Am Soc Nephrol*. 2016;11(11):1935-1943. doi:10.2215/CJN.00280116
10. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. *Crit Care Med*. 2018;46(7):1070-1077. doi:10.1097/CCM.00000000000003123
11. Hodgson LE, Roderick PJ, Venn RM, Yao GL, Dimitrov BD, Forni LG. The ICE-AKI study: impact analysis of a clinical prediction rule and electronic AKI alert in general medical patients. *PLoS One*. 2018;13(8):e0200584. doi:10.1371/journal.pone.0200584
12. Hodgson LE, Sarnowski A, Roderick PJ, Dimitrov BD, Venn RM, Forni LG. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. *BMJ Open*. 2017;7(9):e016591. doi:10.1136/bmjopen-2017-016591
13. Tomašev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572(7767):116-119. doi:10.1038/s41586-019-1390-1
14. Lei VJ, Luong T, Shan E, et al. Risk stratification for postoperative acute kidney injury in major noncardiac surgery using preoperative and intraoperative data. *JAMA Netw Open*. 2019;2(12):e1916921. doi:10.1001/jamanetworkopen.2019.16921
15. ClinicalTrials.gov. An early real-time electronic health record risk algorithm for the prevention and treatment of acute kidney injury. Updated September 27, 2019. Accessed July 10, 2020. <https://clinicaltrials.gov/ct2/show/NCT03590028>
16. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73. doi:10.7326/M14-0698

17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845. doi:[10.2307/2531595](https://doi.org/10.2307/2531595)
18. Murray PT, Mehta RL, Shaw A, et al; ADQI 10 workgroup. Potential use of biomarkers in acute kidney injury: report and summary of recommendations from the 10th Acute Dialysis Quality Initiative consensus conference. *Kidney Int*. 2014;85(3):513-521. doi:[10.1038/ki.2013.374](https://doi.org/10.1038/ki.2013.374)
19. Haase M, Devarajan P, Haase-Fielitz A, et al. The outcome of neutrophil gelatinase-associated lipocalin-positive subclinical acute kidney injury: a multicenter pooled analysis of prospective studies. *J Am Coll Cardiol*. 2011;57(17):1752-1761. doi:[10.1016/j.jacc.2010.11.051](https://doi.org/10.1016/j.jacc.2010.11.051)
20. Nickolas TL, Schmidt-Ott KM, Canetta P, et al. Diagnostic and prognostic stratification in the emergency department using urinary biomarkers of nephron damage: a multicenter prospective cohort study. *J Am Coll Cardiol*. 2012;59(3):246-255. doi:[10.1016/j.jacc.2011.10.854](https://doi.org/10.1016/j.jacc.2011.10.854)
21. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol*. 2015;68(1):25-34. doi:[10.1016/j.jclinepi.2014.09.007](https://doi.org/10.1016/j.jclinepi.2014.09.007)
22. Selby NM, Casula A, Lamming L, et al. An organizational-level program of intervention for AKI: a pragmatic stepped wedge cluster randomized trial. *J Am Soc Nephrol*. 2019;30(3):505-515. doi:[10.1681/ASN.2018090886](https://doi.org/10.1681/ASN.2018090886)
23. Meersch M, Schmidt C, Hoffmeier A, et al. Prevention of cardiac surgery-associated AKI by implementing the KDIGO guidelines in high risk patients identified by biomarkers: the PrevAKI randomized controlled trial. *Intensive Care Med*. 2017;43(11):1551-1561. doi:[10.1007/s00134-016-4670-3](https://doi.org/10.1007/s00134-016-4670-3)
24. Göcze I, Jauch D, Götz M, et al. Biomarker-guided intervention to prevent acute kidney injury after major surgery: the prospective randomized BigpAK Study. *Ann Surg*. 2018;267(6):1013-1020. doi:[10.1097/SLA.0000000000002485](https://doi.org/10.1097/SLA.0000000000002485)
25. Bernier-Jean A, Beaubien-Souligny W, Goupil R, et al. Diagnosis and outcomes of acute kidney injury using surrogate and imputation methods for missing preadmission creatinine values. *BMC Nephrol*. 2017;18(1):141. doi:[10.1186/s12882-017-0552-3](https://doi.org/10.1186/s12882-017-0552-3)

SUPPLEMENT.

eFigure 1. Variable Importance Plot for the Simplified Model Developed in the University of Chicago Derivation Cohort

eFigure 2. Consort Diagram for the 2 External Validation Cohorts

eFigure 3. Cumulative Incidence Plots for Stages of AKI and Receipt of Dialysis Across All 3 Study Cohorts

eFigure 4. Calibration Plots for Stage 2 AKI Across All 3 Study Cohorts

eTable 1. Variables Included in the Gradient Boosted Machine Prediction Model

eTable 2. Characteristics and Outcome Comparisons Between Those With and Without AKI Across Sites

eTable 3. Area Under the Receiver Operating Characteristic Curve for the Model to Predict Stage 2 AKI in the Next 24 Hours in All Cohorts, Stratified by Patient Location, Admission Serum Creatinine Level, and Time in Operating Room

eTable 4. Accuracy and Timing of Detection of Different Probability Cutoffs for Detecting Stage 2 AKI Within the Next 48 Hours Using All Calculated Risk Scores During the Admission Prior to the Event or Discharge